Zuffa Ex. 9

(PCCX685 - William M. Boal & Michael R. Ransom, Monopsony in the Labor Market, Journal of Economic Literature (1997))

Journal of Economic Literature
Vol. XXXV (March 1997), pp. 86–112

Monopsony in the Labor Market

WILLIAM M. BOAL
Drake University
and
MICHAEL R RANSOM
Brigham Young University

We acknowledge helpful comments from David Neumark and several referees.

1. *Introduction*

THE TERM "monopsony," first used by Joan Robinson (1969, p. 215) at the suggestion of her friend, classics scholar B. L. Hallward, literally means a market with a single buyer. Robinson explored the consequences of monopsony in the labor market—in particular, the effects of upward-sloping labor supply to the individual firm—and her simple model is presented in many undergraduate textbooks. Until recently, however, the typical tone of that presentation has been skeptical, as textbook authors and labor economists generally have focused on the implausibility of the single-buyer assumption.

Since Robinson, numerous models of buyer market power have been developed that do not assume a single buyer or even a small number of buyers. Today the term "labor monopsony" is applied more broadly to any model where individual firms face upward-sloping labor supply. New developments on at least three fronts have rekindled labor economists' interest in monopsony under this broader definition. First, new empirical studies have found results that seem to contradict competitive models. These include studies finding apparent positive effects of minimum wages on employment and studies finding an apparent positive effect of firm size on wages that cannot be explained by competitive factors. Second, new theoretical and empirical studies in the job search literature have explored Dale Mortensen's (1970) insight that search behavior induces upward-sloping labor supply to the firm in the short run. Third, recently developed empirical methods from industrial organization, especially those exploiting repeated observations over time, have begun to find their way into labor economics. Meanwhile, a steady stream of empirical research has continued to investigate monopsony using older methods.

Although the literature on labor monopsony draws on the same microeconomic theory as the industrial organization literature on product-market power, much of the labor monopsony literature looks very different from the industrial organization literature for several reasons. Dynamics are more important in the labor market. The large literature on labor market frictions suggests that workers probably switch employers more slowly than most consumers switch brands of products. Also, cross-sectional

data and especially panel data on individual workers are more plentiful than data on individual buyers in product markets. These data allow detailed investigation of supply dynamics. Furthermore, legal minimum wages are much more common in labor markets than are price ceilings in product markets (except for a few industries, like public utilities). Monopsony theories have important (though often ambiguous) implications for the effects of minimum wages. In addition, equity considerations always lurk in the background in the monopsony literature because the wage is such an important determinant of economic welfare, especially for low-income persons. However, the policy prescriptions derived from a simple monopsony model can easily be contraindicated in a more complicated model, as will be seen below. In this survey we therefore emphasize theory and measurement rather than policy and welfare.

2. *Basic Monopsony Models*

The important consequences of upward-sloping supply can be seen clearly in a simple model of a labor market with a single firm. We begin this section therefore with a review of the basic textbook model of a monopsonist and some implications of that model. We then consider simple extensions into dynamic labor supply.

2.1 *The Isolated Firm Model*

Consider a profit-maximizing firm's choice of labor input. Let $L(w)$ denote the firm's labor supply function, where L denotes employment and w denotes the firm's wage. If L is measured in workers, then $L(w)$ is proportional to the cumulative distribution of reservation wages of those workers available to the firm. Under the important special case of perfectly elastic labor supply, this distribution degenerates to a single wage. To accommodate this case, it is more convenient to work with the inverse labor supply function $w(L)$. Let $R(L)$ be the firm's revenue function net of other input costs,[1] with $dR/dL > 0$. The firm's problem then is

$$\max_L R(L) - w(L)\,L \qquad (1)$$

for which the first-order condition is

$$0 = \frac{dR}{dL} - \left(w + \frac{dw}{dL}L\right). \qquad (2)$$

Here, dR/dL is marginal revenue product (MRP), while the expression in parentheses is marginal labor cost (MLC). Their intersection determines monopsony employment L_m and the monopsony wage $w_m = w(L_m)$, as shown in Figure 1 (Robinson 1969, p. 220). If the monopsonist firm could hire all the workers it wanted at wage w_m, it would set employment higher at L_m^*. The difference $L_m^* - L_m$ is sometimes interpreted as the firm's "vacancies" (G. C. Archibald 1954).

The monopsony outcome may be contrasted with the competitive outcome, which is given by the intersection of MRP and labor supply (w_c and L_c in Figure 1). The competitive and monopsony outcomes are identical for this market if $dw/dL = 0$; i.e., labor supply is perfectly elastic, a situation which might be described as zero monopsony power.

The first-order condition (2) can be rearranged to give

$$E \equiv \frac{MRP - w}{w} = \varepsilon^{-1}. \qquad (3)$$

where ε is the elasticity of labor supply. The left-hand side is Arthur Pigou's

[1] If other inputs are variable, it is understood they are set at values that maximize R, given L. If the firm is a nonprofit organization or government, $R(L)$ might be interpreted as some other objective function increasing in L.

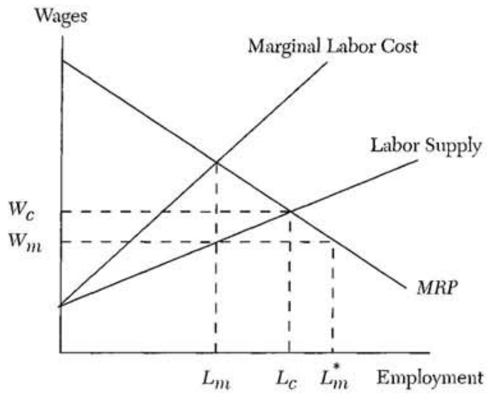

Figure 1. Wage and Employment Determination under Monopsony

(1924, p. 754) measure of "exploitation" and is analogous to the Lerner index often used to measure departures from competition in product markets. Clearly, E measures the departure of wages from marginal revenue product in percentage terms, taking the value zero in the case of competition. It is thus comparable to gaps caused by other labor-market distortions, gaps such as union relative wage effects or marginal tax rates.

E has the virtue of being computable from just the local elasticity of the supply curve. However, E does not measure welfare directly. One welfare measure is deadweight loss relative to the competitive outcome, given by

$$DWL = \int_{L_m}^{L_c} (MRP(L) - w(L))\, dL \qquad (4)$$

but unlike E, deadweight loss is not necessarily decreasing in ε.[2] Nor does E measure the departure of wages from their competitive level (Pigou's "unfairness," 1924, p. 754) unless the *MRP* curve is horizontal. Jointly sufficient conditions for a horizontal (long run) *MRP* curve are (1) constant returns to scale production, (2) perfectly elastic supply of other inputs, and (3) perfectly elastic demand for output (John Hicks 1932, pp. 242–46; Michael Bradfield 1990). The third condition is plausible for private sector firms producing for competitive output markets, but is surely implausible for local public sector employers.

Effects of a minimum wage. Monopsonistic markets respond differently to a minimum wage from competitive markets (George Stigler 1946; Archibald 1954). A minimum below the monopsony wage w_m has no effect, but as the minimum rises above w_m, it creates a kink in the monopsonist's perceived supply curve (and hence a discontinuous *MLC* curve). The profit-maximizing employment and wage are on the supply curve at this kink until the minimum reaches the competitive wage w_c. As the minimum rises further above w_c, the monopsonist's perceived supply curve becomes horizontal over the relevant range, and the optimal employment can be read from the *MRP* curve. Thus, employment determination passes through three regimes as the minimum wage rises: first a nonbinding regime, then a supply-determined regime, and finally a demand-determined regime. These regime shifts create a nonmonotonic relationship between the minimum wage and employment.

The second regime, corresponding to the upward-sloping segment between (w_m, L_m) and (w_c, L_c), where employment is supply-determined, has no counterpart in competitive markets. Along this segment, an increase in the minimum wage increases employment. In particular, the elasticity of employment with respect to the minimum equals the *reciprocal* of E; ironically, policy is most effective when

[2] However, for the special case of horizontal *MRP* and linear labor supply, the ratio of deadweight loss to total competitive earnings ($w_c L_c$) is given by $\frac{1}{4}/(\varepsilon + 1)$, which is decreasing in ε. For constant-elasticity labor supply, the same ratio is decreasing in ε when ε exceeds about 0.3.

E is small! On the other hand, when E is small, the *range* of wages over which a minimum wage tends to boost employment is also small: E gives the maximum percent wage increase that does not decrease employment.

The previous discussion assumes that the firm remains in operation. However, it is conceivable that a minimum wage could decrease profits so much that the firm would shut down before employment ever reaches the competitive level L_c and employment would drop to zero.

Wage discrimination. If the monopsonist firm knows the reservation wages of individual workers, it can engage in first-degree wage discrimination, paying each worker only his reservation wage. In practice, the firm is likely to know at most the elasticities of supply of different groups of workers. If these elasticities vary across groups, the firm can engage in third-degree wage discrimination, setting wages separately for each group. By equation (3), even if all workers are equally productive, groups of workers with higher elasticities of labor supply will enjoy higher wages (Robinson 1969, pp. 224–27).

2.2 *Dynamic Labor Supply*

Now suppose the profit-maximizing firm operates in many periods. Even if short-run labor demand is relatively inelastic, short-run considerations must be balanced against long-run considerations if labor supply responds to wage changes with a lag.

> The possibility of exploitation thus depends on two things: on the ease with which [workers] can move, and on the extent to which they and their employers consider the future, or look only to the moment. (Hicks 1932, p. 83)

This principle can be illustrated by the following very simple model. Suppose the firm's labor supply function function takes the dynamic form: $L_t = L(w_t,\ L_{t-1})$.[3] Because labor supply is likely to adjust slowly to any wage change, both partial derivatives of this function are likely to be positive. Consequently, inverting this function gives:

$$w_t = w_t(L_t, L_{t-1}) \tag{5}$$

with $\partial w_t/\partial L_t \geq 0$, but $\partial w_t/\partial L_{t-1} \leq 0$.[4] If the firm has discount rate r, its problem is now

$$\max_{L_1, L_2, \dots} \sum_{t=1}^{\infty} (R_t(L_t) - w_t(L_t, L_{t-1})\, L_t) \left(\frac{1}{1+r}\right)^{t-1}, \tag{6}$$

for which a representative first-order condition is

$$0 = \frac{dR_t}{dL_t} - w_t - \frac{\partial w_t}{\partial L_t} L_t - \frac{\partial w_{t+1}}{\partial L_t} \frac{L_{t+1}}{1+r}. \tag{7}$$

Assume that the inverse elasticities,

$$\varepsilon_{SR}^{-1} \equiv \frac{\partial w_t}{\partial L_t} \frac{L_t}{w_t}, \quad \varepsilon_1^{-1} \equiv \frac{\partial w_{t+1}}{\partial L_t} \frac{L_t}{w_{t+1}}, \tag{8}$$

are constant over time and that a steady state holds ($L_t = L_{t+1}$ and $w_t = w_{t+1}$).[5] Then the first-order condition can be rearranged to give

$$E_t \equiv \frac{MRP_t - w_t}{w_t} = \varepsilon_{SR}^{-1} + \frac{\varepsilon_1^{-1}}{1+r}. \tag{9}$$

[3] This labor supply specification can be derived from a partial adjustment or adaptive expectations framework (Boal 1995). However, similar results can be obtained from any distributed lag specification.

[4] The signs of these partial derivatives are obtained by totally differentiating $L_t = L(w_t,\ L_{t-1})$ and rearranging the results. An intuitive explanation for the negative derivative associated with L_{t-1} is that the higher the past level of employment at a firm, the lower the wage required to obtain a given level of employment now, because labor supply adjusts slowly.

[5] The simplifying assumption of a steady state rules out cycles of high and low wages. However, the results would be qualitatively similar if wages and employment grew or shrank at constant rates over time.

Finally, the long-run inverse elasticity, derived by equating $L_t = L_{t-1} = L$, is given by $\varepsilon_{LR}^{-1} = (\varepsilon_{SR}^{-1} + \varepsilon_1^{-1})$, so one may write

$$E_t = \varepsilon_{SR}^{-1}\left(\frac{r}{1+r}\right) + \varepsilon_{LR}^{-1}\left(\frac{1}{1+r}\right). \qquad (10)$$

In words, the rate of exploitation is a weighted average of the short-run and long-run inverse elasticities of labor supply, where the weights depend on the discount rate r. If periods are annual or less, then r is likely to be small and the long-run inverse elasticity is weighted much more heavily than the short-run.[6]

Labor supply function (5) includes an interesting special case. Let $h(w_t, L_{t-1})$ express the number of new hires into the firm and $q(w_t, L_{t-1})$ express the number of quits.[7] The current employment level can thus be written as:

$$L_t = L_{t-1} + h(w_t, L_{t-1}) - q(w_t, L_{t-1}) \qquad (11)$$

The short-run inverse elasticity of labor supply, derived by holding L_{t-1} constant, is given by

$$\varepsilon_{SR}^{-1} = \frac{L_t}{w_t}\left(\frac{\partial h}{\partial w_t} - \frac{\partial q}{\partial w_t}\right)^{-1}. \qquad (12)$$

Assuming $\partial h/\partial w_t > 0$ and $\partial q/\partial w_t < 0$, the short-run inverse elasticity is positive. The long-run inverse elasticity, derived by equating $L_t = L_{t-1} = L$, is given by

$$\varepsilon_{LR}^{-1} = \varepsilon_{SR}^{-1}\left(\frac{\partial q}{\partial L} - \frac{\partial h}{\partial L}\right). \qquad (13)$$

Long-run labor supply is less than perfectly elastic only if $\partial q/\partial L > \partial h/\partial L$ and is perfectly elastic if the two partial derivatives are equal. The reason is that under the former condition, larger firms suffer greater net outflows of workers, holding wages constant. To avoid shrinking, larger firms must set higher wages. By contrast, under the latter condition, net outflows of workers are independent of firm size, so larger firms need not set higher wages to avoid shrinking.

This example illustrates the principle that under monopsony with dynamic labor supply, E lies between the short-run and long-run inverse elasticities. This principle implies first, that E is less than the short-run inverse elasticity of labor supply if firms "consider the future" at all. Second, it implies that E can still be positive even if long-run labor supply is perfectly elastic. In the latter case, long-run employment is below the competitive level, though wages are not.[8] In general, the rate of exploitation E depends directly on the firm's discount rate r and inversely on the speed with which labor supply responds to wage changes.

3. *Sources of Monopsony Power*

Textbooks usually interpret monopsony as describing a particular firm with exclusive access to a completely isolated labor market. Such cases are surely rare. When other firms are present, the reservation wages of potential employees depend on the wages offered by other firms. One might expect competition from alternative employers to thwart monopsony, effectively driving the reservation wages of all potential employees up to the competitive wage. What might

[6] Simple dynamics can be introduced on the labor demand side with little change in the results. For example, if the revenue function is given by $R(L_t, L_{t-1})$, then results (9) and (10) still hold provided MRP_t is defined as $\partial R_t/\partial L + (\partial R_{t+1}/\partial L_t)/(1 + r)$.

[7] If the elasticity of hires (quits) with respect to L_{t-1} is unitary, then the hiring *rate* (quit *rate*) is independent of firm size. Some authors discussed below assume the elasticity of quits is indeed unitary, but the elasticity of hires is less than one, perhaps even zero.

[8] If the long-run supply is perfectly elastic at wage $\bar{w}$, then this must be the steady-state wage under either monopsony or competition. Under monopsony, $MRP > \bar{w}$, by equation (10), while under competition, $MRP = \bar{w}$. If MRP is downward-sloping, then L_m must be less than L_c.

prevent this? The literature suggests several possibilities.

3.1 *Oligopsony*

In classic oligopoly models, firms do not take prices as given. Instead, they jointly maximize profits (the collusive model) or take each other's quantities as given (the Cournot model). Although similar models are rarely developed formally in the labor economics literature, many empirical studies comparing wages and employer concentration across labor markets seem to rely implicitly on non-wage-taking models, presumably inspired by similar empirical studies of product markets.[9] A brief review of the collusive and Cournot models will facilitate interpretation of this empirical literature. This subsection maintains the assumption that all workers at all firms receive the same wage, at least in the long run, an assumption relaxed in later subsections.

Adam Smith was convinced that firms

> are always and everywhere in a sort of tacit, but constant and uniform combination, not to raise the wages of labor above their actual rate. (Smith 1937, p. 67)

The collusive model gives the same results as the basic monopsony model, except that "*MRP*" now refers to the horizontal sum of all firms' individual MRP_i curves. Thus, all firms enjoy the same rate of exploitation E.[10] The comparative statics with respect to the number of firms (denoted n) deserves consideration here. On the one hand, if each firm's MRP_i curve is horizontal, then n has no influence on equilibrium E, w, or L, although one might argue that the effectiveness or likelihood of collusion might be inversely related to the number of firms. On the other hand, if each firm's MRP_i curve is downward-sloping, then the addition of a new firm shifts the market MRP curve to the right, increasing both w and L—but this result holds under competition too. In summary, a positive effect of n on w and L is evidence for collusion and against competition only when total market demand for labor by all potentially colluding firms is held constant—a persistent problem in the empirical literature to be discussed below.[11]

Under the Cournot model, firms play an employment-setting game and each firm's problem becomes

$$\max_{L_i} R_i(L_i) - w(L_i + L_i^*)\, L_i, \qquad (14)$$

where L_i is the firm's own employment level, $R_i(L_i)$ is the firm's own revenue function, and L_i^* is the employment level for all other firms in the labor market. A single market wage is determined by the total employment of all firms $L = L_i + L_i^*$. The first-order condition for each firm implies a firm-specific rate of exploitation E_i given by

$$E_i = \frac{MRP_i - w}{w} = \frac{L_i}{L}\,\varepsilon^{-1}, \qquad (15)$$

where ε is again the market-level elasticity of supply. If all firms have identical MRP_i curves (the symmetric case), then they will have identical employment shares (L_i/L) and rates of exploitation E_i. The rate of exploitation will then be inversely proportional to the number of firms, n. If firms have different MRP_i

[9] The studies known to the authors that explicitly develop non-wage-taking models are John Penrod (1995) and Paul Beck (1993). Donald Yett (1970, pp. 379–80) develops a kinked-supply oligopsony model, but, as widely noted in the industrial organization literature, this model has little predictive power, being compatible with any wage-employment outcome from competition to monopsony (Jean Tirole 1988, p. 244).

[10] The first-order conditions for joint profit maximization require that MRP_is for all firms be equated to each other and to (market) MLC.

[11] Total market labor supply may also be correlated with n. If so, then tests of competition are likely to produce false negatives unless labor supply is controlled for.

curves (the asymmetric case), then the same equation shows that firms with high MRP_i curves will have higher market shares and higher rates of exploitation. An employment-weighted average of these rates of exploitation is given by

$$E = \sum_{i=1}^{n} E_i \frac{L_i}{L} = \varepsilon^{-1} \left[\sum_{i=1}^{n} \left(\frac{L_i}{L} \right)^2 \right], \qquad (16)$$

where the sum in brackets is the Herfindahl index of concentration, hereafter denoted "H."

The relationship between E and H is not a comparative static one. Both E and H are endogenous market outcomes and depend on the number of firms and the distribution of MRP_i across them (Harold Demsetz 1973). Nevertheless, a positive correlation across markets between E and H (or E and n) is evidence of non-wage-taking behavior, because if all firms took w as given, the only possible equilibrium implies E equals zero, irrespective of H or n. By contrast, a negative correlation across markets between w and H is not evidence against competition unless labor supply and total market demand for labor are held constant—again, a persistent problem for empirical work.

3.2 *Classic Differentiation*

If firms differ discretely along dimensions of, say, location or working conditions, and workers have heterogeneous preferences on these dimensions, then each firm may enjoy an upward-sloping (inverse) supply function of the form

$$w_i = w_i(L_i, X_i^*), \qquad (17)$$

where X_i^* represents the actions of other firms in the market—e.g., wages (Bertrand) or employment levels (Cournot)—and the firm's problem becomes

$$\max_{L_i} R_i(L_i) - w_i(L_i, X_i^*)\, L_i. \qquad (18)$$

On the one hand, if each worker prefers a particular firm over all others by a least some finite amount, then the only possible Nash equilibrium sets the wage exactly equal to the collusive level (Peter Diamond 1971).[12] On the other hand, if workers' preferences are distributed continuously so that at least some workers are on the margin, then the rate of exploitation is smaller, but not zero. In the latter case, the (inverse) labor supply function (17) is differentiable and upward-sloping because, even though alternative employers are present in the market, their relative attractiveness varies across workers, resulting in a nondegenerate distribution of reservation wages. The first-order condition for each firm implies a firm-specific rate of exploitation E_i equal to that firm's partial elasticity of w_i with respect to L_i.

The comparative statics of classic differentiation with respect to the number of firms is unclear without more structure. However, it seems reasonable that as more firms enter the market, the elasticity of supply to any one firm will grow, and the market will approach hedonic competition.

3.3 *Moving Costs*

When workers must pay costs (whether pecuniary or psychic) to change firms, the resulting model resembles classic differentiation but with a dynamic element. Firms between whom the worker is indifferent at the time of hire become "differentiated" once the employee moves to a particular job location. The importance of this post-hire "differentiation" depends on the length of wage contracts (whether explicit or

[12] This is because at any other wage level, firms will want to cut wages slightly, because a wage cut smaller than workers' utility differential will not cause any workers to leave. Here, the inverse labor supply function is vertical over a small range.

implicit). If the firm can commit to wages for the entire length of an employee's tenure, a competitive market results. If the firm cannot commit to wages, the firm enjoys monopsony power when wages come up for renegotiation.[13] As firms exploit this power, and workers anticipate exploitation, the wage-seniority profile becomes front-loaded, showing downward slope—or at least less upward slope than in competitive markets (Dan Black and Mark Loewenstein 1991).

Moreover, if moving costs differ across workers and are not observed by the firm, then the firm confronts a nondegenerate distribution of reservation wages. The firm must inevitably drive away some low-moving-cost employees in order to exploit the rest, so employment (of long-tenured workers) will be reduced relative to a competitive market. Alternatively, if the firm observes differences in mean moving costs by group, the wage-discriminating firm will offer the lowest ex post wage to the group with the highest moving costs. These predictions are very similar to those of the textbook monopsony and discriminating monopsony, respectively, but moving-cost models also offer predictions about turnover. Turnover will always be increased relative to the competitive outcome in that market. In equilibrium, low-moving-cost workers will move frequently but command high wages—perhaps even higher than competitive wages (Yannis Ioannides and Christopher Pissarides 1985; Black and Loewenstein 1991; Ransom 1993).

Post-hire exploitation is possible only if moving costs to workers are important, turnover costs to firms are unimportant, and firms cannot commit to future wages. Therefore, this model seems unlikely to apply to unskilled workers, whose alternative employers are in close proximity and whose wages can be roughly specified in advance at low cost. It also seems unlikely to apply to managers or other skilled workers with substantial specific training and therefore high turnover costs to firms. It seems most likely to apply to professionals with general skills whose alternative employers are geographically dispersed and whose wages cannot be specified far in advance—such as college professors (Black and Loewenstein 1991; Ransom 1993).

3.4 *Equilibrium Search*

Models of job search assume that it takes time for workers and firms to find each other. Thus a firm's flow of new hires is bounded by a finite flow of job applicants, and the inverse elasticity of labor supply to the firm must be positive in the short run—see equation (12). However, some recent search models imply a positive long-run inverse elasticity as well. These models are called "equilibrium search" models because their main motivation is to explain the distribution of offered wages as the outcome of optimizing behavior by both workers and firms. A searching worker's reservation wage is optimal only if at least one firm actually offers that wage, while conversely an employer's offered wage is optimal only if it is the reservation wage of at least one worker. Thus the set of wages actually offered by firms expecting to attract workers (formally, the support of the offered-wage distribution) must be identical to the set of reservation wages of workers expecting to find jobs (the support of the reservation-wage distribution). How can this set of wages show dispersion in equilibrium?

One strand of literature, beginning

[13] Formally, absence of precommitment implies that labor contracts must be "self-enforcing" or equivalently that the equilibrium of the game played between workers and firms must be "subgame-perfect."

with James Albrecht and Bo Axell (1984), derives wage dispersion by assuming exogenous heterogeneity. Workers, initially unemployed, are assumed to have varying values of nonmarket time while firms are assumed to have varying productivity levels. In equilibrium, more productive firms offer higher wages, and workers who value nonmarket time have high reservation wages. Nevertheless, all employed workers are paid less than their marginal product.

Another strand of literature, beginning with Kenneth Burdett and Mortensen (1989), derives wage dispersion by allowing employed workers to search for jobs. The reservation wages of employed workers are simply their current wages. This is sufficient to guarantee wage dispersion in equilibrium, even though firms and workers are homogeneous ex ante.

The basic model of Burdett and Mortensen has just five exogenous parameters: b, the value of nonmarket time enjoyed by unemployed workers; MRP, the (constant) marginal revenue product of employed workers; λ_0, the arrival rate of job offers to unemployed workers; λ_1, the arrival rate of job offers to employed workers; and δ, the exogenous rate at which worker-firm matches break up. (Here, breakups will be interpreted as worker exits from the labor force, but they may also be interpreted as job destruction.) Discounting does not occur, and firms are assumed to care only about the long run. When firms play a one-shot wage posting game, it can be shown that the equilibrium distribution of wage offers is nondegenerate and has compact support $[w_L, w_H]$. In particular, this distribution takes the form

$$F(w) = \left(1 + \frac{\delta}{\lambda_1}\right)\left(1 - \sqrt{\frac{MRP - w}{MRP - w_L}}\right), \quad w \in [w_L, w_H]. \quad (19)$$

This wage distribution is easily shown to be stochastically increasing either in w_L or in the ratio (λ_1/δ), holding the other quantity constant.

The highest and lowest wages are, of course, endogenous. The lowest wage, w_L, the reservation wage of unemployed workers is determined in the market as follows. If wage offers arrive no faster for employed workers than for unemployed workers ($\lambda_1 \leq \lambda_0$), then w_L equals b, the value of nonmarket time. If offers arrive faster for employed workers ($\lambda_1 > \lambda_0$), then a job provides not only a wage but a means to a better job, so w_L is less than b. However, if a legal minimum wage greater than the market-determined lowest wage is imposed, then w_L becomes the legal minimum. The highest wage, w_H, found by setting $F(w_H) = 1$ in equation (19), is easily shown to equal the following weighted average of w_L and MRP:

$$w_H = w_L\left(\frac{\delta}{\delta + \lambda_1}\right)^2 + MRP\left(1 - \left(\frac{\delta}{\delta + \lambda_1}\right)^2\right). \quad (20)$$

w_H is increasing in the ratio (λ_1/δ), but nevertheless $w_H < MRP$ as long as δ is positive.

Different wage offers in the support of F yield different levels of steady-state employment. What makes equation (19) an equilibrium distribution is that all wage offers can be shown to yield the same steady-state *profits*; i.e., an equal-profit condition $\pi = (MRP - w)L$ holds for all firms. This last condition can be interpreted as the long-run supply of labor to any individual firm because firms are homogeneous. Totally differentiating this condition with respect to w and L and rearranging terms gives the familiar result

$$\varepsilon_{LR}^{-1} = \frac{MRP - w}{w} > 0. \quad (21)$$

Because wages are positively related to firm size, the distribution of firms over wage offers $F(w)$ is different from the distribution of workers over wages. It can be shown that the latter distribution is given by

$$G(w) = \frac{F(w)}{1 + (\lambda_1/\delta)(1 - F(w))}. \quad (22)$$

What prevents a firm from expanding employment without raising its wage in this model? The answer lies in the assumed matching technology, termed "random matching" in the literature, which assumes that *each firm is equally likely to make an offer to a given worker*. Put differently, workers sample from the distribution $F(w)$, not $G(w)$. Firms consequently suffer diseconomies of scale in hiring workers. To see this, let m denote the measure of employed workers in the labor market, let u denote the measure of unemployed workers, and let n denote the measure of firms (all continua). Hires are obtained through offers to unemployed workers or to employed workers currently earning lower wages, so the flow of new hires is given by

$$h(w,L) = \frac{1}{n}(\lambda_0 u + \lambda_1 m G(w)). \quad (23)$$

Quits are exogenous (δ) or induced by offers from better-paying firms, so the flow of quits is given by

$$q(w,L) = (\delta + \lambda_1[1 - F(w)])\, L. \quad (24)$$

Equations (23) and (24) show that quits are proportional to firm size but hires are not. Thus $\partial_q/\partial_L > \partial_h/\partial_L$ and the long-run inverse elasticity, derived by equating $h(w,L) = q(w,L)$, is positive; see equation (13).

Alternative matching technologies can give different results, of course. For example, one simple alternative technology is "balanced matching," wherein workers sample from $G(w)$, not $F(w)$. Under "balanced matching," hires are proportional to firm size, and it can be shown that the equilibrium distribution of offered wages is degenerate at the competitive wage $w = MRP$ (Burdett and Tara Vishwanath 1988). However, if random and balanced matching coexist, some wage dispersion and monopsony can still be supported if a sufficient fraction of search is of the random kind (Mortensen and Vishwanath 1994). Some writers identify "balanced matching" empirically with worker search through personal contacts and "random matching" with search through publicly advertised vacancies or gate applications, but such an identification probably makes too much of an extremely stylized model.[14] The essence of the Burdett-Mortensen model is that in search models, diseconomies of scale in the net hiring function $h(w,L) - q(w,L)$ can support steady-state equilibrium wage dispersion and monopsony, even with (ex ante) homogeneous workers and firms and without time discounting.

3.5 *Efficiency Wages at Large Firms*

Consider an efficiency-wage model in which firms economize on monitoring costs by paying above-equilibrium wages. If firms suffer diseconomies of scale in monitoring workers, as James Rebitzer and Lowell Taylor (1995) assume and Guillermo Calvo and Stanislaw Wellisz (1979) derive from a hierarchical model, then when the firm increases employment, it must increase wages to maintain the required penalty for shirking. The result is upward-sloping supply in the long run, implying that the wage must be below marginal revenue product (the dif-

[14] Some evidence on hiring rates and firm size is given in Charles Brown and James Medoff (1989, pp. 1048–49). Comparatively little attention has been given to the realism of the quit function, though evidence cited in the same source (pp. 1041–44) and elsewhere suggests that quit rates decline with firm size.

ference is the increase in efficiency wages for inframarginal workers) and employment is reduced below the competitive perfect-information level. In contrast to search models, however, imperfect information here *increases* wages above their competitive level! While this model is intuitively plausible, the empirical evidence for diseconomies of scale in monitoring is only mixed (C. Brown and J. Medoff 1989, pp. 1051–55; Francis Green, Stephen Machin, and Alan Manning 1992, pp. 13–15).

3.6 *Comparisons and Contrasts*

All the models presented in this section share with the isolated firm model the following features: (1) labor supply to the individual firm is upward-sloping, at least in the short run; (2) the firm sets its wage below the marginal product of labor; (3) and the firm sets its employment level below the competitive level. However, the models are quite different in other respects, such as the following.

Implied size of the market: The isolated-firm, collusive, and Cournot models are most naturally interpreted as representing a particular labor market within a larger economy, for two reasons. First, they assume a small number of firms. Second, they assume upward-sloping labor supply to the market—more workers are drawn into employment as the wage rises. Because labor supply to the entire economy is nearly vertical, most of these additional workers must be drawn from other labor markets. By contrast, equilibrium search models are more naturally interpreted as representing an entire economy, or at least the labor market for an entire demographic group of workers, because they assume a large number of firms and a fixed pool of labor. In between are models based on classic differentiation, moving costs, and efficiency wages, which can be interpreted as representing either a particular labor market or the entire economy.

Concentration and exploitation: Concentration is related to the rate of exploitation directly in the Cournot model—see equation (16). This relationship also exists roughly in the collusive model, to the degree that concentration facilitates coordination, and possibly in the models of classic differentiation and moving costs, to the degree that concentration proxies for distances between firms and moving costs. In contrast, concentration plays no role in efficiency-wage and equilibrium-search models.

Firm size and exploitation: In the asymmetric Cournot model, E_i is *positively* related to the firm's employment level L_i. In the basic search model of Burdett and Mortensen, by contrast, all firms lie on the same labor supply curve, so that by equation (21), E_i is *negatively* related to L_i.[15] In the remaining models, the relation between E_i and L_i cannot be inferred without more assumptions.

Turnover and exploitation: Typical textbook presentations of the isolated-firm model stress the firm's physical separation, suggesting that monopsony power must be negatively related to employee turnover. By contrast, this suggestion is misleading or incorrect for moving-cost and search models, in which turnover is more or less endogenous. In moving-cost models, turnover and exploitation are *negatively* correlated across workers with varying moving costs in the same market, as mentioned above. Across markets, however, the correlation can be positive—holding average moving costs constant, those markets that permit wage commitment avoid turnover and pay all workers the competitive wage, while those markets without commit-

[15] In this respect, search models echo Pigou's (1924, p. 534) view that "the small masters, throughout history, have always been the worst exploiters."

ment suffer inefficient turnover and pay low wages to immobile workers. In search models, turnover and exploitation are *positively* correlated across firms in the same market, because equation (24) implies that $q(w,L)/L$ is inversely related to w.[16] Across markets, however, the correlation is not as clear cut: for example, equations (19) and (22) show that the distributions of wages $F(w)$ or $G(w)$ are stochastically decreasing in the exogenous breakup rate δ, holding λ_0 and λ_1 constant, but are unaffected by a proportionate change in all three parameters.

4. *What Can Monopsony Explain?*

Monopsony models of the labor market are inherently more complicated than competitive models. What does this complication buy? What qualitative features of the labor market can monopsony models explain easily that competitive models explain only awkwardly?

4.1 *Vacancies*

Under competition, firms can hire all the workers they want at the going wage. That firms sometimes report they cannot is therefore a puzzle for competitive models. Under monopsony, persistent vacancies—interpreted as measuring excess demand assuming *MRP* slopes downward—are easy to explain. Excess demand persists because labor supply to the firm slopes upward (Archibald 1954). For example, Yett (1970, pp. 371–75) cites vacancies as evidence of monopsony in the U.S. market for nurses in the late 1960s. Nevertheless, how much stock should be put in firms' reported desires, as opposed to their actions, is open to debate (Sherwin Rosen 1970, pp. 391–92).

[16] Evidence that turnover rates are inversely related to firm size and establishment size, respectively, are given in Mary Miner (1977, p. 30) and John Pencavel (1970, p. 59).

4.2 *Persistent Wage Dispersion*

Under competition, equally productive workers at equally attractive jobs should earn identical wages. An awkward fact is that often workers that appear identical to researchers are paid different wages. (See for example William Dickens and Lawrence Katz 1987.) But wages in monopsonized markets need not equal wages in other markets. Moreover, wage dispersion within markets is a central feature of equilibrium search models.

4.3 *The Employer Size-Wage Relationship*

Numerous studies document significant correlations between wages and the size of the firm or establishment that are difficult for competitive models to explain (C. Brown and J. Medoff 1989; Green, Machin, and Manning 1992). Some monopsony models cannot explain them either. For example, under the isolated firm model or classic differentiation, size and wages are positively correlated across firms only if size differences are driven mostly by shifts in *MRP* curves, rather than shifts in supply curves. Under collusive and Cournot models, size and wages should be *negatively* correlated across markets, ceteris paribus, assuming that average firm size is positively correlated with concentration.

However, efficiency-wage and search models predict positive correlations between size and wages. Indeed, the inverse elasticity of size with respect to wages (with appropriate controls—see Section 5.3 below) is exactly the rate of exploitation. Under efficiency-wage models, wages should most closely relate to *firm* size, assuming monitoring problems are a firm-level phenomenon. Under search models, wages should most closely relate to *establishment* size, as-

suming that workers search over establishments.

4.4 *Effects of Minimum Wages*

A number of recent studies find that increases in the legal minimum wage have no effect, or possibly a positive effect, on aggregate employment (see David Card and Allan Krueger, 1995, for a survey). A positive effect, in particular, contradicts competitive models but is compatible with monopsony. (However, even in a competitive labor market, an increase in the minimum wage may increase employment at *particular firms*.)[17]

The collusive, Cournot, and classic differentiation models all predict a positive effect, provided that the minimum wage is still below *MRP*. However, in the classic differentiation model, the market's labor supply is less elastic than each firm's labor supply, because an increase in wages at other firms—X_i^* in equation (17)—shifts each firm's labor supply curve leftward. The employment effect of the minimum wage is governed by *market* labor supply, while the rate of exploitation E depends on each *firm's* labor supply. It follows that the (local) employment elasticity with respect to the minimum wage will be less than $1/E$, while the maximum percent wage increase that does not decrease employment is greater than E. Finally, the moving-cost model predicts a reduction in turnover, as post-hire exploitation is constrained by the minimum wage.

Equilibrium search models can sometimes predict a positive effect of minimum wages. In these models, an increase in the minimum wage can influence total employment through two possible effects: first an increase in the rate at which workers exit unemployment; and second, a decrease in the number of available jobs. Neither effect occurs in the basic homogeneous model of Burdett and Mortensen (1989), provided the minimum wage is still below *MRP*, because unemployed workers always accept every job offer they receive. However, the first effect can occur with heterogeneous workers because then some workers will have reservations wages above the lowest offered wage, w_L. Their exit rate from unemployment can be increased by a binding minimum wage. The second effect can occur with heterogeneity in *MRP* in particular because a minimum wage that benefits high-*MRP* workers can put low-*MRP* workers out of work. Both effects can occur in a fully heterogeneous model. In that case, a rising minimum wage typically first increases and then decreases employment, just as in the isolated firm model (Burdett and Mortensen 1989; Zvi Eckstein and Kenneth Wolpin 1990; Manning 1994).[18]

There is some evidence that an increase in the minimum wage tends to increase the wages of workers above the new minimum, i.e., to shift up the entire wage distribution (Jean Grossman 1983). While this can occur in competitive models if workers at different wage levels are heterogeneous and gross substitutes, it always occurs in equilibrium search models even if workers are homogeneous (Burdett and Mortensen 1989).

4.5 *Wage Discrimination*

Many studies find that women, African-Americans, Hispanics, and other groups earn less than white men in the U.S. labor market, even after controlling for observed productivity differentials.

[17] Suppose firms are heterogeneous with regard to technology. Firms employing relatively few low skilled workers may actually expand output and employment if product prices rise sufficiently (see Walter Oi 1983, pp. 76–77).

[18] The studies cited all assume the output price is fixed.

These differentials can be explained by (third degree) monopsony wage discrimination if the labor supply of these groups to individual firms can be shown to be less elastic than the labor supply of white men. The literature has given some attention to the case of women. Empirical studies usually find that women's labor supply is more elastic than men's at the level of the market, but some researchers argue informally that this relationship reverses at the level of the individual firm. For example, this might occur under collusive or Cournot models if markets for women's labor were more highly concentrated on the employer side. It might occur under search models if women suffered from a higher exogenous quit rate (δ) and lower offer arrival rates (λ_0, λ_1) (Manning 1993).

There is also some indirect empirical evidence. Manning (1993) shows that both wages and employment of women increased in the United Kingdom with the implementation of the Equal Pay Act; this might be interpreted as minimum wage legislation for women. Green, Machin, and Manning (1992) show that the employer size-wage relationship is stronger for women than for men in the U.K. There is evidence that pay disparities between men and women are strongest in small or highly concentrated labor markets (Robert Frank 1978; Rudolf Winter-Ebmer 1995).

5. *Measuring the Rate of Exploitation*

The case for monopsony in labor markets seems almost compelling. The basic idea of upward-sloping supply, at least in the short run, surely fits the experiences of employers as they attempt to set wage policies. Moreover, monopsony models can explain some features of labor markets not easily explained by competitive models. But do actual labor markets deviate substantially enough from competition to justify abandoning competitive models for more complicated and less tractable monopsony models?

To answer this question, one would first like to know the size of E, the rate of exploitation, in monopsonized markets. Is it as large as, say, union relative wage effects or marginal tax rates, distortions that also drive wedges between labor supply and labor demand? Second, how widespread is monopsony? Is the labor market as a whole or large parts of it characterized by sizeable values for E? Third, for policy purposes, are the positive employment effects of minimum wages in the isolated firm model likely to carry over to the real world (recall the equivocal implications of search models with heterogeneity, for example)? In other words, are instances of low wages in the economy primarily due to monopsony or to low productivity? This section surveys various approaches to measuring E, interprets the results reported in the literature, and briefly discusses the empirical literature on effects of minimum wages.

5.1 *Direct Measurement of Wage and MRP*

All the models presented in Section 2 above imply *MRP* exceeds w, assuming firms maximize profits. An omnibus test for monopsony power therefore compares estimated values of the *MRP*, perhaps from a production function, with actual wages. In principle, this approach should detect the presence of monopsony power, though it would not identify its source. In practice, this approach encounters the typical problems of estimating production functions: measurement of inputs and outputs, functional form questions, endogeneity of inputs, and unobserved inputs correlated with observed inputs. The measurement of outputs may be particularly difficult in public or pri-

vate service sectors (e.g., education, nursing). Moreover, the wage (or rather total compensation) must be measured accurately in levels. A mere wage index is useless. In particular, the measured wage must include benefits and the worker's share of any investment in human capital.[19]

This strategy for measuring monopsony has been used extensively in analyzing the market for professional athletes. Professional sports leagues in North America are organized with distinctive monopsonistic characteristics, such as the "draft" for new players, and agreements between teams that restrict mobility of players within the league.

Almost all the studies examine professional baseball. The interest in baseball is motivated partly by the explicit monopsonistic organization of the league. Until 1976, all players were bound to individual teams by the "reserve clause," which prohibited teams from competing for players.[20] Baseball lends itself to empirical analysis because detailed information on pay and performance of individual players is readily available. Also, baseball is much more individualistic than other sports, so it is not difficult to isolate the effect of one player's performance from that of other members of the team.

The most influential paper on the topic is by Scully (1974). Almost everyone who has published on the topic has adopted some version of his approach. Scully tests monopsony by estimating the *MRP* of individual players and comparing it to pay, so the crux of the method is estimation of *MRP*.

The essential assumption of Scully's model is that a team's revenues increase when the team wins more games, and the performance of players contributes to team revenues *only* by changing the team's winning percentage. *MRP* is estimated via a two-step process. In the first step, team revenues are regressed against winning percentage, along with other team-specific factors, including the population of the metropolitan area in which the team resides, a measure of "fan interest" in the area, whether the team belongs to the National League, an indicator for "old" stadiums, and the fraction of the team's players that are black. Using data from the 1968 and 1969 seasons, Scully estimates that an increase of one percentage point in the team's winning percentage will increase revenues by $10,330.

Scully simplifies by assuming that non-pitchers contribute to winning percentage only through hitting, and pitchers contribute only through pitching. The relationship is estimated by regressing the team's winning percentage against the team's "slugging average" and the team's ratio of strike-outs to walks. Scully also includes variables to measure whether a team was "in the cellar" or "in contention."[21] He finds that a one point increase in the team slugging average increased winning percentage by .92, and a 1/100 point increase in the strike-out-to-walks ratio increased winning percentage by .90.

[19] In human capital models, the nominal wage in any given period need not equal the *MRP* in that period, even under competition.

[20] The reserve clause essentially bound a player to a single team indefinitely, or until the team transferred the player to another team. Various forms of the reserve clause were in effect in major league baseball until 1976, when it was essentially eliminated as a result of an arbitration ruling between the owners and the players' union. Currently, players who have been in the major leagues for at least six years are free to offer their services to other teams. Players with three to six years of service are eligible for final offer arbitration, if the player and team to which he is bound cannot agree on a salary. Gerald Scully (1989) provides details.

[21] Such variables are clearly not exogenous to the winning percentage of the team, so it is difficult to interpret the coefficient on the performance variables.

Thus the estimated marginal product of a hitter in Scully's model is .92 times $10,330 times the player's slugging average times the player's fraction of the team's at-bats. *MRP* for pitchers is defined analogously. Comparing actual salaries with estimated *MRP*, Scully finds that average players were paid about 20 percent of estimated *MRP*, and "star" players were paid about 15 percent of *MRP*. The corresponding values of *E* lie between 4 and 7.

Other authors apply slightly different versions of Scully's model to data from the reserve clause era. For example, Marshall Medoff (1976) estimated that players were paid 30 to 50 percent of *MRP*—a value of *E* between 1 and 2. Andrew Zimbalist (1992) found that exploitation fell after the demise of the reserve clause—he estimated an average value for *E* only slightly greater than zero for 1989, but higher for earlier years. Even after the demise of the reserve clause, there may have been some collusion between teams.

Scully's model is easy to criticize. The premise that spectators are willing to pay only to see performance that contributes to winning by the home team, is tenuous. It is true that baseball teams with higher winning percentages generate more revenues, if the size of the market city is held constant.[22] However, it is also true that teams sell more tickets when a good team visits. Also, fans pay to see athletic performances even when it is not clear that there is a "home team," as in track meets or ice skating competitions. Perhaps the causality is wrong—fine performances attract spectators but contribute incidentally to winning. A direct measure of how individual performance contributes to revenues is needed.

The National Football League provides an interesting illustration. In the NFL, teams share gate revenues much more evenly, and television revenues are shared equally by all teams. Because each team's revenues cannot vary much with winning percentage, the *MRP* (as defined in Scully's model) will vary little with winning percentage. It is not surprising that when Atkinson, Stanley, and Tschirhart (1988) applied Scully's definition of *MRP* to the NFL, they found that the "offensive units" of teams in the National Football League were paid significantly *more* than *MRP*, even though football players do not have free agency. This suggests that Scully's concept of *MRP* is inappropriate.

Even if the premise of the model is correct, the method still suffers from naive modeling of the "production function" for baseball teams. If other inputs into the production of wins are correlated with hitting and pitching performance, then Scully's approach will overstate the *MRP* of players. Also, these studies fail to model carefully either the effect of human capital investments by the teams or the risk involved in developing talent. Scully's notion of *MRP* also fails to recognize that baseball rosters are limited. He assumes that if a star hitter were not on the team, no one would bat in his place. With fixed roster size, a player's net *MRP* is the difference between his contribution and the contribution of the next-best player available.

Scully's (1974, 1989) econometric specifications have also been criticized. Small changes in the model result in large differences in estimated *MRP*. For example, MacDonald and Reynolds (1994, footnotes 9 and 12 in particular) report that crucial coefficients change by about 25 percent in different specifications.

[22] This continues to be true in baseball (Zimbalist 1992; Don MacDonald and Morgan Reynolds 1994). It is also true in the National Football League (Scott Atkinson, Linda Stanley, and John Tschirhart 1988) and the National Hockey League (John Jones and William Walsh 1987).

Incidentally, the most successful cartel in all of sports is not a professional league, but the National Collegiate Athletic Association (NCAA). The NCAA strictly limits the amount that member colleges and universities can pay to their players. (See Gary Becker, 1985, for an interesting discussion.) Robert Brown (1993) estimates that the annual *MRP* of a premium college football player is approximately $500,000. However, this calculation is subject to most of the same criticisms we have made of Scully.

The strategy of directly comparing *MRP* and wages has also been used to analyze the market for U.S. coal miners in the early twentieth century. Isolated mining towns are frequently cited in textbooks as examples of monopsony. Nonunion coal towns in early twentieth century Appalachia enjoy special notoriety because the local coal companies often controlled the police, and typically functioned as landlords, retail merchants, and creditors for their employees. Lawrence Boyd (1994) estimated the *MRP* of coal miners in two West Virginia counties in the early twentieth century, using mine-level data on coal production. Estimates from different years gave widely varying *MRP* estimates which in some cases are sharply less than actual wages. Samples pooled across years gave estimated values of E of 0.24 for one county, but the estimate was not significantly different from zero.

5.2 *Cross-sectional Comparison of Wages and Employer Concentration*

Models that attribute monopsony power to collusion or Cournot behavior may imply an inverse relationship across markets between the wage level and measures of labor market concentration, holding constant the market-level labor supply and *MRP* curves. A large literature has therefore estimated market-level equations of the form

$$w = f(C, X, Y), \tag{25}$$

where C measures employer concentration, X represents determinants of labor supply, and Y represents determinants of *MRP*. Cross-sectional data are used, so the estimated wage-concentration relationship is taken to be a long-run one.

This specification is similar to that used in a vast literature on product markets, beginning with Joe Bain (1951) and surveyed by Richard Schmalensee (1989). Generally speaking, the difficulties encountered in using this approach in labor markets are those encountered in product markets, plus a few more. In particular, it should be noted that the concentration-wage approach can detect monopsony power only if labor supply is less than perfectly elastic in the long run, for otherwise under any monopsony model employment might be reduced but the wage would be unaffected (see Section 2.2 above).

Concentration could be measured in various ways: as a four-firm or eight-firm concentration ratio, as a Herfindahl index, or as the total number of firms. Any of these measures may be interpreted in a model of joint profit maximization as measuring the likelihood of successful coordination in a collusive model. The Herfindahl index can be interpreted in a Cournot model as the factor relating E to the inverse elasticity of market labor supply. Alternatively, these concentration measures might be interpreted as gauging the degree of differentiation among firms, but perhaps geographic density of employers might better represent lack of differentiation. Thus, ignoring efficiency-wage and search models, the wage-concentration relationship would seem to be an omnibus test for employer market power, provided labor

supply is not perfectly elastic in the long run.

No concentration measure can be computed without defining the relevant market. Most studies of labor markets use metropolitan areas or counties, arguing that worker mobility is limited to a commuting radius (Robert Bunting 1962; James Luizer and Robert Thornton 1986). However, this market definition may be too small, given the frequency of worker relocation in the United States and recalling that the mobility of only the *marginal* worker is relevant.

Although in principle, concentration is an outcome of the model, and thus is not exogenous, one could argue that at least in public sector labor markets concentration is largely determined by predetermined political boundaries. However, C might still be correlated with unobserved elements of X and Y. For example, X should include wages in related labor markets and the cost of living, variables that tend to be closely correlated with population and population density in the labor market, but which in turn tend to be closely correlated with C. Similarly, Y should include determinants of MRP, but the danger of omitted MRP shifters is, if anything, at least as large as for supply shifters. For example, a labor market with more employers (and therefore lower concentration) is likely to be one with greater demand, both for final output and for workers.

Several researchers have compared market concentration and wages to test for monopsony power in the market for public school teachers. School districts are candidates as monopsonists because of their natural geographical separation. Some have argued that school teachers are also more susceptible to exploitation because a high proportion are married women who are "tied" to their husbands' jobs, limiting their mobility. Luizer and Thornton (1986) and Beck (1993) survey studies of the market for public school teachers.

A typical study is Luizer and Thornton (1986), which analyzes data from 266 school districts in 15 metropolitan areas of Pennsylvania for the 1978–79 school year. Districts are included in the "labor market" of an area if most of the nonresident teachers of the district lived within 15 miles of the central city of the metropolitan area. Luizer and Thornton estimate regression models to explain the scheduled salaries of teachers at several levels of education and experience, using various indexes to measure employer concentration, including the Herfindahl index and the one-firm and four-firm concentration ratios based on the number of teachers employed by each district. Variables used to identify shifts in labor supply and demand for teachers are number of students in the district, personal income per student of district residents, property tax rate, and the percentage of the district's population living in urban areas.

Luizer and Thornton find statistically significant monopsony effects only for teachers with bachelor degrees at five or ten years of experience. For a teacher with five years experience and a bachelor degree, their estimates predict that moving from the most concentrated market to the least concentrated market would increase salary by $400 to $500, depending on the concentration index used—approximately three percent of the average teacher salary for Pennsylvania in that year ($15,200).

Beck's (1993) dissertation is the most comprehensive of the studies of monopsony in the school-teacher market. He analyzes pooled data from all 541 school districts in Missouri for several years between 1982 and 1990. He defines the market for each district to include all districts located within a 25 mile radius. As a dependent variable, he uses the

logarithm of the average teacher's salary in a district, and as explanatory variables the average experience and education characteristics of the district's teachers, along with measures of the district's size, urban nature, racial composition, tax base, income of its residents, sex composition, and union membership of the district's teachers, and dummy variables for years. The Herfindahl index based on number of teachers employed measures employer concentration.

Beck finds a small monopsony effect in his analysis of all districts—splitting two districts into four equal-sized districts in a market would increase salaries by slightly more than one percent. However, he finds quite different results for urban and rural districts. In rural districts, where monopsony might be thought of as more likely, higher concentration is actually associated with higher salaries.

Applications of this empirical strategy to the market for nurses are also numerous, and sometimes studies find large impacts on salaries. The study by Charles Link and John Landon (1976) is typical. It analyzes survey data from 317 hospitals in 1973. The dependent variable is the hospital's starting salary for a registered nurse with no previous nursing experience. Salaries are explained in a regression model using variables that indicate the type of control (private nonprofit, private for-profit, Veteran's Administration, other), average hourly wage in manufacturing in the city, a price index for the city, indicators of nonwage benefits associated with the hospital, and a measure of hospital concentration—the "entropy" of the number of hospital beds in the city.[23] Link and Landon find a large, statistically significant effect of concentration. The authors estimate that the annual salary of a registered nurse in Lynchburg, Virginia (the most concentrated city in their study) would increase by $1,600 per year if concentration were to fall to the level of New York City (the least concentrated market in their study). Link and Landon do not report sample averages, but average salary from another survey of nurse wages for the same year is about $8,000, so $1600 represents approximately 20 percent of the national average salary. This is after alleged accounting for differences in cost-of-living (by including the mean wage of manufacturing employees and a cost-of-living index for each city as explanatory variables). Most of the other published studies of nurses' salaries and hospital concentration find some support for monopsony in the market for hospital nurses.

Unfortunately, concentration in the hospital market is very closely correlated with the size of the urban area—no large metropolitan areas have only a few hospitals; no small towns have many hospitals. The same is true for school districts, although in this case political rather than economic boundaries are more important. Because housing and commuting costs will be much higher in large urban centers, it is not surprising that those who work in such communities are paid more than those who work in small cities or towns.

The body of evidence from studies comparing concentration with wages fails to be convincing, because so many variables that potentially could change the *MRP* or the supply curve are missing. The most damaging evidence in this respect comes from Roger Feldman and Richard Scheffler (1982), who find that the effect of hospital concentration on the wages of hospital housekeepers is almost exactly the same as for registered

[23] Entropy is defined as $\sum s_i \log(s_i)$, where s_i is the share of the ith hospital of all hospital beds and the sum is taken over all hospitals in the city. This study apparently uses the boundaries of a city as the boundaries of the labor market area of the hospitals in that city.

nurses. It seems unlikely that concentration of hospitals would result in monopsonization of the market for unskilled labor. Furthermore, Killard Adamache and Frank Sloan (1982) find that when population density of the market is included in regression models, the apparent effect of employer concentration disappears. Penrod (1995) reports the same phenomenon when analyzing salaries of college and university professors. Barry Hirsch and Edward Schumacher (1995) also fail to find an effect of concentration on nurses' hourly earnings in Current Population Survey data. They compare nursing wages to wages of similar workers in the Metropolitan Statistical Area (MSA) or state (for nonmetropolitan areas). Their approach probably does a better job of controlling for differences in cost-of-living between markets.

5.3 *Estimation of Elasticity of Labor Supply to Individual Firm*

All models except collusion require that (perceived) labor supply to an individual firm be less than perfectly elastic, at least in the short run. A number of studies attempt to measure monopsony power by estimating the perceived inverse labor supply function to an individual firm. Assuming the individual firm maximizes profits, the value of the inverse elasticity of labor supply provides an estimate of E_i. Perhaps the most obvious approach is to estimate some form of equation (17), relating wages to employment using firm-level data.[24] Panel data are preferable to cross-sections because they allow the researcher to control for firm heterogeneity. Just as important, panel data facilitate estimation of labor-supply dynamics, which recent studies have suggested are critical.[25]

In a widely cited paper, Sullivan (1989) estimates equation (17) for the supply of nurses to individual hospitals. Sullivan's data consist of nursing wages and employment levels at several thousand U.S. hospitals observed over six years. To explore dynamics informally, Sullivan differences the data at one-year, two-year, and three-year intervals before estimation. Supply is substantially more elastic in the long run than in the short run, the estimated inverse elasticities descending from about 0.75 at one-year intervals to about 0.26 at three-year intervals. However, the rate of exploitation E_i is probably less than either of these figures. The long-run inverse elasticity cannot be computed without a formal model of dynamics, but it is plausible to suppose that it is close to zero.[26] Suppose it is zero, and suppose further that Sullivan's one-year estimate of 0.75 represents the short-run inverse elasticity and that the dynamic model of equation (5) applies. By equation (10), E_i would equal 0.04 if hospitals' discount rate is 5 percent, about 0.07 if the discount rate is 10 percent, and about 0.13 if the discount rate is 20 percent. While Sullivan shows that his estimates are robust to the

[24] Instrumental variables estimation is presumably necessary because employment and wages are determined simultaneously. The studies described below, by Daniel Sullivan (1989), Korinna Hansen (1992), and Boal (1995), use exogenous measures of output demand as instruments.

[25] If changes in employment are negatively correlated with average tenure at a firm, and tenure is positively correlated with wages, then estimates of the supply elasticity will be biased upwards, at least for the short run. Sullivan (1989, pp. 2159–60) acknowledges this problem. Boal (1995) avoids it by using piece-rate wages.

[26] If the true labor supply function is a loglinear version of equation (5): $\ln(w_t) = \beta_1 + \beta_2\ln(L_t) + \beta_3\ln(L_{t-1})$, then it can be shown that the short-run inverse elasticity is β_2, the one-period inverse elasticity is $\beta_2/(1 - \beta_3/\beta_2)$, and the two-period inverse elasticity is $\beta_2/(1 - \beta_3/\beta_2 + [\beta_3/\beta_2]^2)$; see Boal 1995, p. 524. Fitting these expressions to all of Sullivan's full-sample estimates of one-year, two-year, and three-year inverse elasticities, respectively, by nonlinear least squares gives $\beta_2 = 0.754$ and $\beta_3 = -0.766$, implying a long-run inverse elasticity of $\beta_2 + \beta_3 = -0.012$.

choice of specification, there is some evidence that they are sensitive to the choice of data set. Hansen (1992) estimates nearly identical specifications on a California data set (Sullivan's data set included the entire U.S.) but finds one-year inverse elasticities generally less than 0.05.

Boal (1995) applies a similar econometric method to West Virginia coal-mining data from the early twentieth century. Estimating a log-linear version of equation (5), Boal computes short-run inverse elasticities of 0.15 to 0.53 depending on specification. However, long-run inverse elasticities are essentially zero, implying that E_i are at most 0.03, 0.05, or 0.09, using discount rates of 5 percent, 10 percent, or 20 percent, respectively. Moreover, Boal believes these estimates are biased upward because they are based on county-level, rather than firm-level or mine-level data.

Several practical issues regarding the specification of the covariates that represent other firms' actions (X_i^*) have not been cleanly resolved in these panel studies. The first issue is whether to include other firms' wages (Bertrand) or employment levels (Cournot) or some other actions. It is surely possible to devise a formal test for the appropriate equilibrium concept (perhaps using estimates of *MRP*), but the power of any such test is likely to be low, and none has yet been reported. Studies to date (Sullivan 1989; Hansen 1992; Boal 1995) have been able to sidestep this question because their results were not very sensitive to it, but future studies may not be so lucky. The second issue is how to aggregate other firms' actions. If a given firm interacts with many others in the same market, the effect of each firm may be impossible to estimate separately. Other firms' actions must first be aggregated. But how many other firms should be included? Including too many or too few could bias the results. The old antitrust question of appropriate market size reappears in this new econometric setting! Perhaps the most sensible approach, at least for exploratory work, is to introduce separate variables for (aggregated) nearby firms and (aggregated) farther-away firms and let their relative coefficients determine the appropriate market size.[27]

Cross-sectional data sets on firms are an attractive alternative to time-series or panel data sets because they typically contain more observations with greater variation in L. Unfortunately, controlling for firm heterogeneity and dynamics is more difficult in a cross-sectional framework. A simple dynamic structure can be estimated if data on employment flows (hires and quits) or the previous period's employment level are available, but otherwise, estimates of the long-run inverse elasticity are likely to be biased upward slightly (because large firms are more likely to have grown recently).[28] One control that may *not* be needed in cross-sectional studies is X_i^*—the actions of other firms. Omission of X_i^* is surely invalid when using time-series or panel data. (Indeed, Boal, 1993 found it changed the results noticeably.) However, it may be valid when using cross-sectional data for a single labor market

[27] This approach is used by Boal (1993), who finds that nearby firms do indeed exert greater influence than distant firms. However, the estimates of the own inverse elasticities are not substantially affected in his case by allowing separate effects.

[28] For example, suppose a log-linear version of equation (5) were estimated omitting lagged employment. Application of equation (9) and standard omitted-variable analysis shows that the expectation of the estimated coefficient of current employment would equal not the long-run elasticity $\varepsilon_{SR}^{-1} + \varepsilon_1^{-1}$ but rather $\varepsilon_{SR}^{-1} + \rho\varepsilon_1^{-1}$, where ρ is the serial correlation in employment. Jonathan Leonard (1987, p. 153, table 6.7) estimates ρ to be about 0.97. Estimates of ε_1^{-1} given by Boal (1995) and implied by Sullivan (1989) range from 0.2 to 0.8. (Boal, 1995, p. 524.)

under the assumption that the individual firm is atomistic, for then X_i^* is essentially the same for all firms in the market whether the solution concept is Bertrand, Cournot, or some other symmetric notion. This atomistic assumption is standard for search models, although probably less reasonable for models based on geographic differentiation.

Machin, Manning, and Stephen Woodland (1993) estimate E_i using a cross-sectional sample of residential homes for the elderly in England. The estimate of ε_{LR}^{-1}, obtained from least-squares regressions of wages on employment, is roughly 0.04, an estimate typical of other data sets (C. Brown and J. Medoff 1989, pp. 1034–35, 1038, 1040).[29] The paper's distinctive feature is its attempt to correct this value for differences in unobserved worker quality and in hedonic job amenities or disamenities associated with firm size. Addressing the former, the authors, finding that output price (a measure of worker quality after controlling for other inputs) is positively associated with firm size, conclude that worker quality is positively associated with firm size, requiring a small downward correction. Addressing the latter, the authors find that quits rise less than proportionately with firm size after controlling for wages, and conclude that *positive* job amenities exist at larger firms (a typical finding but a debatable conclusion). The authors propose a positive correction calculated as the percentage decrease in wages required to keep the quit propensity constant as employment rises by one percent. This latter correction turns out to be substantial—about 0.19 at the sample mean of employment. After further adjustment for short-run supply response (using data on quits and hires) and assuming a discount rate of 5 percent, the authors arrive at estimates of E_i for each firm in their sample. The mean estimated E_i is 0.15, but presumably would be much smaller without the authors' correction for size-related amenities.

5.4 *Wages and Mobility*

Ransom (1993) estimates indirectly the rate of exploitation of college professors in the United States. In his model, employers enjoy monopsony power due to geographical dispersion, so employees must incur a cost to move from one employer to another. The national market for professors is competitive—a university must pay the "market wage" to hire a professor. However, once a professor accepts a contract and moves to the location of a specific university, he or she must pay moving costs to leave for another university. Thus each year universities hire labor in two different markets, one of which is competitive (the external market), and the other monopsonized (the internal market).

The internal market is monopsonistic because the employer can keep a higher fraction of current employees by offering a higher wage in contract renewals. However, the employer minimizes costs by offering current employees less than the "market wage." In Ransom's model, workers are further differentiated by costs of moving. In equilibrium, those with the highest moving costs will be paid less and will move less often. Thus salaries will fall with seniority. After controlling for total teaching experience, education level, and other productivity-related characteristics, Ransom finds maximum "seniority penalties" of about 5 to 15 percent in three national surveys of university faculty, and for faculty at the University of Arizona.

According to Ransom's model, movers are paid the market wage, which might

[29] Whether ordinary least-squares regressions of wages on employment estimate ε_{LR}^{-1} consistently is debatable. If employment is determined simultaneously with wages, presumably the OLS estimate of ε_{LR}^{-1} is biased downward.

be thought of as *MRP*. Thus, the seniority penalty of long-tenure employees measures their exploitation. Ransom's seniority penalties correspond to values of E of 5 to 18 percent. However, as Black and Loewenstein (1991) point out, movers may be paid more than *MRP*, because universities expect to exploit them in the future. Therefore, Ransom's estimates should be considered upper bounds.

5.5 *Structural Estimation Using Equilibrium Search Model*

The equilibrium search model of Burdett and Mortensen (1989) implies monopsony as a consequence of diseconomies of scale in hiring workers. The average rate of exploitation depends inversely on the ratio (λ_1/δ), i.e., on the rate at which employed workers receive offers from other employers relative to the rate at which job matches exogenously break up. Burdett and Mortensen's model is extremely powerful, predicting the entire wage distribution over firms or workers, transition rates from unemployment to employment and from one employer to another, the unemployment rate of workers, and the employer size-wage relationship, all from just a few parameters. These parameters are therefore easily overidentified in typical panel data sets on workers.

However, an obvious difficulty for this model concerns the shape of the wage distribution. Differentiation of (18) or (21) shows that the density of wages over either firms or workers is increasing in w in this model and therefore is skewed extremely negatively. Because actual wage distributions tend to be positively skewed, empirical implementations must include heterogeneity or measurement error (or both) to have any hope of fitting actual wage data. Moreover, ignoring heterogeneity and measurement error is likely to bias upward the estimated average rate of exploitation. This is because wages are dispersed below *MRP* in this model. Any wage dispersion not attributed to heterogeneity or measurement error is effectively attributed to monopsony.

Two recent papers estimate Burdett and Mortensen's model, applying maximum likelihood methods to panel data sets on workers and including heterogeneity to different degrees. Nicholas Kiefer and George Neumann (1993) estimate the model using U.S. data, permitting the parameters to vary by race and education level. Gerard van den Berg and Geert Ridder (1993) estimate the model using data from The Netherlands, permitting the parameters to vary by occupation category, age, and education level, assuming measurement error on wages, and permitting additional unobserved heterogeneity in the *MRP* parameter. Allowing for greater heterogeneity apparently yields lower estimated average rates of exploitation: van den Berg and Ridder's estimates of average E are roughly 0.13 to 0.15, whereas Kiefer and Neuman's estimates are three to five times higher.[30] Nevertheless, both sets of estimates are much larger than employer size-wage effects estimated in earlier research; these measure the same quantity, according to equation (21). For example, estimates in C. Brown and J. Medoff (1989, pp. 1304–05) imply an average E of roughly 0.01 to 0.03. One suspects that even the method of van den Berg and Ridder does not permit sufficient heterogeneity, and that the true rate of exploitation lies between 0.03 and 0.13.

[30] The differences in estimates might also be due to differences in the data sets, of course. Average rates of exploitation reported here are calculated by inserting estimated values for *MRP* and the expected value of w over firms (i.e., over $F(w)$) into equation (3). Using instead the expected value over workers (i.e., over $G(w)$) would yield a slightly smaller rate of exploitation.

The apparent importance of heterogeneity in *MRP* implies that a nontrivial minimum wage is likely to eliminate jobs. This is because a minimum wage set high enough to raise wages for a significant fraction of workers is likely to exceed the *MRP* of some of them. Indeed, van den Berg and Ridder (1993, p. 24) conclude that a 25 percent increase in the existing minimum wage would exceed the *MRPs* of 16 percent of the individuals in their sample. Eckstein and Wolpin (1990, p. 805) arrive at similar conclusions using a different model.

5.6 *Employment Effects of Minimum Wages*

A huge literature investigates the impact of minimum wage increases on employment at both firm and aggregate levels. Early studies (reviewed by C. Brown, Curtis Gilroy, and Andrew Kohen 1982) using aggregate U.S. data through the 1970s generally estimate negative elasticities of employment (of teenagers and sometimes young adults) with respect to the minimum wage (hereafter denoted ε_{MW}) of about 0.1 to 0.3. Later studies using data from the 1980s are more varied. Some estimate ε_{MW} to be in the same range as earlier studies (David Neumark and William Wascher 1995), others estimate ε_{MW} to be essentially zero (Alison Wellington 1991), while still others estimate ε_{MW} to be small but positive (see Card and Krueger 1995).

Unfortunately, estimates of ε_{MW} are nearly useless for testing monopsony or measuring the rate of exploitation E. Recall that, except in the isolated firm, collusive, and Cournot models, ε_{MW} is smaller than the elasticity of supply to individual firms, for two reasons raised earlier. First, an increase in the minimum wage increases the wage offered by other firms. Second, if there is significant heterogeneity in productivity across workers or firms, some workers are likely to become unemployed as the minimum wage rises, even at low levels. Thus, negative estimates of ε_{MW} do not refute monopsony; and positive estimates, after taking reciprocals, are likely to be much greater than E.

Of course, estimates of ε_{MW} are of interest in their own right as a guide to policy. But even if the labor market were entirely monopsonized, ε_{MW} is likely to vary and even to change sign with the level of the minimum wage. One recent minimum-wage study permits ε_{MW} to change signs, following the regime changes of the isolated firm model (see Section 2.2). Neumark and Wascher (1994) estimate competitive and monopsony models of the labor market in a switching regression framework using U.S. state-level data. The competitive model has two regimes: a nonbinding regime and a labor-demand-curve regime. The monopsony model has three regimes: a nonbinding regime, a supply-curve regime, and a demand-curve regime (see Section 2.2 above). The two models are not nested, but nonnested tests seem to favor the monopsony model slightly. The estimated switch points, expressed relative to the average wage for workers, are w_c = 0.35 for the competitive model and w_m = 0.31 and w_c = 0.34. These values are slightly greater than the historic low values of the U.S. federal minimum wage before it was increased in 1990 and 1991. A note of caution: the regime-shifting structure used by Neumark and Wascher is elegant but a bit misleading when applied to aggregate data. The elasticity of their supply-curve regime is not the elasticity of supply to any firm (which cannot be inferred from aggregate data, for the reasons given above) but just the maximum value of ε_{MW}. Nevertheless, flexible specification of ε_{MW} is surely the right empirical approach when the data show substantial variation in the minimum

wage (as in the U.S. over the last few decades), whether the true model is competition or some form of monopsony.

6. *Conclusion*

Monopsonistic exploitation arising from supply frictions, whether modeled as differentiation or search, is probably widespread but small on average. Dynamic studies of workers and firms suggest that short-run inverse elasticities of supply to many firms are probably large. However, long-run inverse elasticities are probably no higher than previous estimates of the elasticity of wages with respect to firm size, about 0.03 or 0.04. At sensible discount rates, the rate of exploitation is likely to be only a little higher than these latter values. Even studies of textbook examples of monopsony such as nursing and coal mining suggest rates of exploitation no higher than about 0.15 and sometimes as low as zero. Thus rates of exploitation arising from supply frictions are probably lower than, say, union relative wage effects or marginal income tax rates faced by U.S. workers.

Monopsonistic exploitation arising from explicit collusion is probably rare but occasionally large. Well-documented cases include U.S. baseball before the reserve clause and perhaps other professional sports. Even in these cases, however, the best estimates of the rate of exploitation reported to date are probably not very accurate.

Monopsonistic exploitation arising from tacitly collusive or Cournot behavior may exist in some professions with small numbers of employers, but the existing evidence is inconclusive. Studies comparing employer concentration and pay most often find no effect when sufficient controls are included. This negative result suggests an absence of market power, but it could also reflect perfectly elastic long-run supply, in which case monopsony depresses employment but not wages. (See Section 2.2, above.)

Under monopsony models more sophisticated (and arguably more realistic) than the usual textbook model of an isolated firm, the effect of minimum wages can be complicated. On the one hand, even if rates of exploitation are small on average, legal minimum wages may still raise (at least some) wages substantially without reducing employment. On the other hand, even if the entire labor market is monopsonized, minimum wages are ineffective when wage dispersion is primarily due to heterogeneity in marginal products rather than heterogeneity in rates of exploitation. Thus, under sophisticated models, the effect of minimum wages on employment remains an empirical question. Conversely, monopsony models are not easily refuted by empirical studies of the employment effect of minimum wages.

REFERENCES

ADAMACHE, KILLARD W. AND SLOAN, FRANK A. "Unions and Hospitals: Some Unresolved Issues," *J. Health Econ.*, May 1982, *1*(1), pp. 81–108.

ALBRECHT, JAMES W. AND AXELL, BO. "An Equilibrium Model of Search Unemployment," *J. Polit. Econ.*, Oct. 1984, *92*(5), pp. 824–40.

ARCHIBALD, G. C. "The Factor Gap and the Level of Wages," *Econ. Record*, Nov. 1954, *30*(59), pp. 187–99.

→ ATKINSON, SCOTT E.; STANLEY, LINDA R. AND TSCHIRHART, JOHN. "Revenue Sharing As an Incentive in an Agency Problem: An Example from the National Football League," *RAND J. Econ.*, Spring 1988, *19*(1), pp. 27–43.

→ BAIN, JOE S. "Relation of Profit Rate to Industry Concentration: American Manufacturing, 1936–1940," *Quart. J. Econ.*, Aug. 1951, *65*(3), pp. 293–324.

BECK, PAUL M. "Monopsony in the Market for Public School Teachers in Missouri: The Static and Dynamic Impact On Salaries and Employment." Ph.D. Dissertation, U. of Missouri, 1993.

BECKER, GARY S. "College Athletes Should Get Paid What They Are Worth," *Business Week*, Sept. 30, 1985, (No. 2914), p. 18.

VAN DEN BERG, GERARD J. AND RIDDER, GEERT. "An Empirical Equilibrium Search Model of the

Labour Market." Vrije Universiteit, Amsterdam, Faculty of Economics and Econometrics Research Memorandum 1993–39, July 1993.

BLACK, DAN A. AND LOEWENSTEIN, MARK A. "Self-enforcing Labor Contracts with Costly Mobility: The Subgame Perfect Solution to the Chairman's Problem," *Res. Lab. Econ.*, 1991, *12*, pp. 63–83.

BOAL, WILLIAM M. "Testing for Employer Monopsony in Turn-Of-The-Century Coal Mining." Unpub. paper. Ohio State U., Sept. 1993.

———. "Testing for Employer Monopsony in Turn-of-the-Century Coal Mining," *RAND J. Econ.*, Autumn 1995, *26*(3), pp. 519–36.

BOYD, LAWRENCE W. "The Company Town, the Union, and the Mine Wars: the Role of Institutions and Conflict in West Virginia's Early Mine Labor Market." Paper presented at the annual meetings of the Economic History Association, Cincinnati, Oct. 1994.

BRADFIELD, MICHAEL. "Long-run Equilibrium under Pure Monopsony," *Can. J. Econ.*, Aug. 1990, *23*(3), pp. 700–04.

BROWN, CHARLES; GILROY, CURTIS AND KOHEN, ANDREW. "The Effect of the Minimum Wage On Employment and Unemployment," *J. Econ. Lit.*, June 1982, *20*(2), pp. 487–528.

BROWN, CHARLES AND MEDOFF, JAMES. "The Employer Size-Wage Effect," *J. Polit. Econ.*, Oct. 1989, *97*(5), pp. 1027–59.

BROWN, ROBERT W. "An Estimate of the Rent Generated by a Premium College Football Player," *Econ. Inquiry*, Oct. 1993, *31*(4), pp. 671–84.

BUNTING, ROBERT L. *Employer concentration in local labor markets*. Chapel Hill: U. of North Carolina Press, 1962.

BURDETT, KENNETH AND MORTENSEN, DALE T. "Equilibrium Wage Differentials and Employer Size." Northwestern Center for Mathematical Studies in Economics and Management Science Working Paper 860, Oct. 1989.

BURDETT, KENNETH AND VISHWANATH, TARA. "Balanced Matching and Labor Market Equilibrium," *J. Polit. Econ.*, Oct. 1988, *96*(5), pp. 1048–65.

CALVO, GUILLERMO A. AND WELLISZ, STANISLAW. "Hierarchy, Ability, and Income Distribution," *J. Polit. Econ.*, Oct. 1979, *87*(5), pp. 991–1010.

CARD, DAVID E. AND KRUEGER, ALAN B. *Myth and measurement: The new economics of the minimum wage*. Princeton, New Jersey: Princeton U. Press, 1995.

DEMSETZ, HAROLD. "Industry Structure, Market Rivalry, and Public Policy," *J. Law Econ.*, Apr. 1973, *16*(1), pp. 1–9.

DIAMOND, PETER A. "A Model of Price Adjustment," *J. Econ. Theory*, June 1971, *3*(2), pp. 156–68.

DICKENS, WILLIAM T. AND KATZ, LAWRENCE F. "Inter-Industry Wage Differences and Industry Characteristics," *Unemployment and the structure of labor markets*. in Eds.: KEVIN LANG AND JONATHAN S. LEONARD. New York: Basil Blackwell, 1987, pp. 48–89.

ECKSTEIN, ZVI AND WOLPIN, KENNETH I. "Estimating a Market Equilibrium Search Model from Panel Data on Individuals," *Econometrica*, July 1990, *58*(4), pp. 783–808.

FELDMAN, ROGER AND SCHEFFLER, RICHARD. "The Union Impact on Hospital Wages and Fringe Benefits," *Ind. Lab. Relat. Rev.*, Jan. 1982, *35*(2), pp. 196–206.

FRANK, ROBERT H. "Why Women Earn Less: The Theory and Estimation of Differential Overqualification," *Amer. Econ. Rev.*, June 1978, *68*(3), pp. 360–73.

GREEN, FRANCIS; MACHIN, STEPHEN AND MANNING, ALAN. "The Employer Size-Wage Effect: Is Monopsony the Explanation?" Centre for Economic Performance Discussion Paper No. 79, June 1992.

GROSSMAN, JEAN BALDWIN. "The Impact of the Minimum Wage On Other Wages," *J. Human Res.*, Summer 1983, *18*(3), pp. 359–78.

HANSEN, KORINNA. "Registered and Practical Nurses in California: Competition Versus Oligopsony in the 1980s." Working paper, Oct. 1992.

HICKS, JOHN R. *The theory of wages*. London: Macmillan, 1932.

HIRSCH, BARRY T. AND SCHUMACHER, EDWARD J. "Monopsony Power and Relative Wages in the Labor Market for Nurses," *J. Health Econ.*, Oct. 1995, *14*(4), pp. 443–76..

IOANNIDES, YANNIS M. AND PISSARIDES, CHRISTOPHER A. "Monopsony and the Lifetime Relation between Wages and Productivity," *J. Lab. Econ.*, Jan. 1985, *3*(1), pp. 91–100.

JONES, JOHN C. H. AND WALSH, WILLIAM D. "The World Hockey Association and Player Exploitation in the National Hockey League," *Quart. Rev. Econ. Bus.*, Summer 1987, *27*(2), pp. 87–101.

KIEFER, NICHOLAS M. AND NEUMANN, GEORGE R. "Wage Dispersion with Homogeneity: The Empirical Equilibrium Search Model," in *Panel data and labour market dynamics*. Eds.: HENNING BUNZEL, PETER JENSEN, AND NIELS WESTERGÅRD-NIELSEN. Amsterdam: North-Holland, 1993, pp. 57–64.

LEONARD, JONATHAN S. "In the Wrong Place at the Wrong Time: The Extent of Frictional and Structural Unemployment," in *Unemployment and the structure of labor markets*. Eds.: KEVIN LANG AND JONATHAN S. LEONARD. New York: Basil Blackwell, 1987, pp. 141–63.

LINK, CHARLES R. AND LANDON, JOHN H. "Market Structure, Nonpecuniary Factors and Professional Salaries: Registered Nurses," *J. Econ. Bus.*, Winter 1976, *28*(2), pp. 151–55.

LUIZER, JAMES AND THORNTON, ROBERT. "Concentration in the Labor Market for Public School Teachers," *Ind. Lab. Relat. Rev.*, July 1986, *39*(4), pp. 573–84.

MACDONALD, DON N. AND REYNOLDS, MORGAN O. "Are Baseball Players Paid Their Marginal

Products?" *Manage. Dec. Econ.*, Sept.-Oct. 1994, *15*(5), pp. 443–57.

MACHIN, STEPHEN; MANNING, ALAN AND WOODLAND, STEPHEN. "Are Workers Paid Their Marginal Product? Evidence from a Low Wage Labor Market." Working paper, Oct. 1993.

MANNING, ALAN. "The Equal Pay Act as an Experiment to Test Theories of the Labour Market." LSE Centre for Economic Performance Discussion Paper No. 153, June 1993.

———. "Labour Markets With Company Wage Policies." LSE Centre for Economic Performance Discussion Paper No. 214, Dec. 1994.

MEDOFF, MARSHALL H. "On Monopsonistic Exploitation in Professional Baseball," *Quart. Rev. Econ. Bus.*, Summer 1976, *16*(2), pp. 113–21.

MINER, MARY G. "Job Absence and Turnover: A New Source of Data," *Mon. Lab. Rev.*, Oct. 1977, *10*(10), pp. 24–31.

MORTENSEN, DALE T. "A Theory of Wage and Employment Dynamics," in *Microeconomic foundations of employment and inflation theory*. Ed.: EDMUND S. PHELPS. New York: Norton, 1970, pp. 167–211.

MORTENSEN, DALE T. AND VISHWANATH, TARA. "Personal Contacts and Earnings: It Is Who You Know!" *Labour Econ.*, Mar. 1994, *1*(2), pp. 187–201.

NEUMARK, DAVID AND WASCHER, WILLIAM. "Minimum Wage Effects and Low-Wage Labor Markets: A Disequilibrium Approach." NBER Working Paper No. 4617, Jan. 1994.

———. "The Effect of New Jersey's Minimum Wage Increase on Fast-Food Employment: A Reevaluation Using Payroll Records." NBER Working Paper No. 5224, Aug. 1995.

OI, WALTER Y. "The Fixed Employment Costs of Specialized Labor," in *The measurement of labor cost*. Ed.: JACK TRIPLETT. Chicago: U. of Chicago Press, 1983, pp. 63–116.

PENCAVEL, JOHN H. "An Analysis of the Quit Rate in American Manufacturing Industry." Industrial Relations Section Working Paper, Princeton U., 1970.

PENROD, JOHN R. "A Test for Monopsony in the Academic Labor Market." Working paper, U. of Michigan, Sept. 1995.

PIGOU, ARTHUR C. *The economics of welfare*. 2nd ed., London: Macmillan and Co., 1924.

RANSOM, MICHAEL R. "Seniority and Monopsony in the Academic Labor Market," *Amer. Econ. Rev.*, Mar. 1993, *83*(1), pp. 221–33.

REBITZER, JAMES B. AND TAYLOR, LOWELL J. "The Consequences of Minimum Wage Laws: Some New Theoretical Ideas," *J. Pub. Econ.*, Feb. 1995, *56*(2), pp. 245–55

ROBINSON, JOAN. *The economics of imperfect competition*. 2nd ed. London: Macmillan, 1969.

ROSEN, SHERWIN. "Comment On Donald E. Yett, "The Chronic 'Shortage' of Nurses: A Public Policy Dilemma," in *Empirical studies in health economics: Proceedings of the second conference on the economics of health*. Eds.: HERBERT E. KLARMAN AND HELEN H. JASZI. Baltimore: Johns Hopkins Press, 1970, pp. 390–94.

SCHMALENSEE, RICHARD. "Inter-industry Studies of Structure and Performance," in *Handbook of industrial organization*. Eds.: RICHARD SCHMALENSEE AND ROBERT D. WILLIG. Amsterdam: North-Holland, 1989, pp. 951–1009.

SCULLY, GERALD W. "Pay and Performance in Major League Baseball," *Amer. Econ. Rev.*, Dec. 1974, *64*(6), pp. 915–30.

———. *The business of major league baseball*. Chicago: U. of Chicago Press, 1989.

SMITH, ADAM. *An inquiry into the nature and causes of the wealth of nations*. New York: Modern Library, [1776] 1937.

STIGLER, GEORGE J. "The Economics of Minimum Wage Legislation," *Amer. Econ. Rev.*, June 1946, *36*, pp. 358–65.

SULLIVAN, DANIEL. "Monopsony Power in the Market for Nurses," *J. Law Econ.*, Oct. 1989, *32*(2, Part 2), pp. S135–78.

TIROLE, JEAN. *The theory of industrial organization*. Cambridge, MA: MIT Press, 1988.

→ WELLINGTON, ALISON J. "Effects of the Minimum Wage on the Employment Status of Youths: An Update," *J. Human Res.*, Winter 1991, *26*(1), pp. 27–46.

WINTER-EBMER, RUDOLF. "Sex Discrimination and Competition in Product and Labour Markets," *Appl. Econ.*, Sept. 1995, *27*(9), pp. 849–57.

YETT, DONALD E. "The Chronic 'Shortage' of Nurses: A Public Policy Dilemma," in *Empirical studies in health economics: Proceedings of the second conference on the economics of health*. Eds.: HERBERT E. KLARMAN AND HELEN H. JASZI. Baltimore: Johns Hopkins Press, 1970, pp. 357–89.

ZIMBALIST, ANDREW. "Salaries and Performance: Beyond the Scully Model," in *Diamonds are forever: The business of baseball*. Ed.: PAUL M. SOMMERS. . Washington, DC: Brookings Institution, 1992, pp. 109–33.